Quant II

TSCS Data Estimation

Ye Wang

3/25/2020

Outline

- What is unique about TSCS data?
- Inference without unobservable confounders
- Inference with unobservable confounders
  - Model-based inference
  - Trajectory-based inference

What is unique about TSCS data?

- Terminology: TSCS or panel data?

What is unique about TSCS data?

- Terminology: TSCS or panel data?
- Why not just pooled OLS?

What is unique about TSCS data?

- Terminology: TSCS or panel data?
- Why not just pooled OLS?
- 1. Observations are dependent; 2. The extra dimension allows us to relax the identification assumptions.

What is unique about TSCS data?

- Terminology: TSCS or panel data?
- Why not just pooled OLS?
- 1. Observations are dependent; 2. The extra dimension allows us to relax the identification assumptions.
- In the cross-sectional setting, we need ignorability: $D_i \perp Y_i(D_i) | X_i$, $0 < P(D_i = 1 | X_i) < 1$.

What is unique about TSCS data?

- Terminology: TSCS or panel data?
- Why not just pooled OLS?
- 1. Observations are dependent; 2. The extra dimension allows us to relax the identification assumptions.
- In the cross-sectional setting, we need ignorability: $D_i \perp Y_i(D_i)|X_i$, $0 < P(D_i = 1|X_i) < 1$.
- Now we can relax the assumption along two possible directions (based on your belief!):
  - Sequential ignorability: $D_{it} \perp Y_{it}(D_{it}, \ldots, D_{i1})|Y_{i,t-1}, D_{i,t-1}, X_{it}$.
  - "Fixed effects:" $D_{it} \perp Y_{it}(D_{it})|X_{it}, \alpha_i, \xi_t$

Which assumption should you choose?

- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.

Which assumption should you choose?

- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history. Sequential experiment, design-based

Which assumption should you choose?

- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history. Sequential experiment, design-based
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.

Which assumption should you choose?

- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.
  Sequential experiment, design-based
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.
  Outcome models (FEct, gsynth, mc, etc.), model-based

Which assumption should you choose?

- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.
  Sequential experiment, design-based
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.
  Outcome models (FEct, gsynth, mc, etc.), model-based
- What if you believe in neither?

Which assumption should you choose?

- Assumption I: All the paths between unobservables and the outcome could be blocked by conditioning on the unit's history.
  Sequential experiment, design-based
- Assumption II: The unobservables have a low-dimension approximation and history doesn't matter.
  Outcome models (FEct, gsynth, mc, etc.), model-based
- What if you believe in neither?
  Make parametric assumptions

Sequential ignorability: then and now

- The assumption is a weaker version of "select on observables."
- All the techniques we have learned are applicable (regression, weighting, matching).
- A big name we cannot ignore in this literature:

Sequential ignorability: then and now

- The assumption is a weaker version of "select on observables."
- All the techniques we have learned are applicable (regression, weighting, matching).
- A big name we cannot ignore in this literature:
  Neal!

Sequential ignorability: then and now

- The assumption is a weaker version of "select on observables."
- All the techniques we have learned are applicable (regression, weighting, matching).
- A big name we cannot ignore in this literature:
  Neal!
- Old idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.) to control for confounders.

Sequential ignorability: then and now

- The assumption is a weaker version of "select on observables."
- All the techniques we have learned are applicable (regression, weighting, matching).
- A big name we cannot ignore in this literature:
  Neal!
- Old idea: we write down a model for Y_{it} (AR, MA, ARIMA, ADL, etc.) to control for confounders.
  The lagged depedent variable does solve a lot of problems, as the time trend is usually the most important confounder.

Sequential ignorability: then and now

- Modern perspective: dynamic treatment regime (popular in epidemiology)

Sequential ignorability: then and now

- Modern perspective: dynamic treatment regime (popular in epidemiology)
- Suppose you participate in a trial to compare two treatments

Sequential ignorability: then and now

- Modern perspective: dynamic treatment regime (popular in epidemiology)
- Suppose you participate in a trial to compare two treatments
  You get treatment I, but it doesn't help you much

Sequential ignorability: then and now

- Modern perspective: dynamic treatment regime (popular in epidemiology)
- Suppose you participate in a trial to compare two treatments
  You get treatment I, but it doesn't help you much
  Several weeks later, the doctor moves you to treatment II

Sequential ignorability: then and now

- Modern perspective: dynamic treatment regime (popular in epidemiology)
- Suppose you participate in a trial to compare two treatments
  You get treatment I, but it doesn't help you much
  Several weeks later, the doctor moves you to treatment II
  How to estimate the effect of your treatment history?

Sequential experiments: estimation

- Suppose there are T weeks, then there are 2^T possible "potential histories."

Sequential experiments: estimation

- Suppose there are T weeks, then there are 2^T possible "potential histories."Too many!

Sequential experiments: estimation

- Suppose there are T weeks, then there are 2^T possible "potential histories."Too many!
- We need to define a meaningful estimand.

Sequential experiments: estimation

- Suppose there are T weeks, then there are 2^T possible "potential histories." Too many!
- We need to define a meaningful estimand.
- Usually we construct a function to aggregate the history's effect: $\tau_i = g(\mathbf{D_i^T}) - g(\mathbf{D_i^{T'}})$.

Sequential experiments: estimation

- Suppose there are T weeks, then there are 2^T possible "potential histories."Too many!
- We need to define a meaningful estimand.
- Usually we construct a function to aggregate the history's effect: $\tau_i = g(\mathbf{D_i^T}) - g(\mathbf{D_i^{T'}})$.
- For example, the number of periods under the treatment

Sequential experiments: estimation

- Suppose there are T weeks, then there are 2^T possible "potential histories."Too many!
- We need to define a meaningful estimand.
- Usually we construct a function to aggregate the history's effect: $\tau_i = g(\mathbf{D_i^T}) - g(\mathbf{D_i^{T'}})$.
- For example, the number of periods under the treatment
- We can obtain valid estimate for the history's aggregated effect after controlling for confounders.

Sequential experiments: estimation

- Yet there is a problem

Sequential experiments: estimation

- Yet there is a problem
- If we control $Y_{i,t-1}$, what will happen?

Sequential experiments: estimation

- Yet there is a problem
- If we control $Y_{i,t-1}$, what will happen?
  Post-treatment bias for $D_{i,t-1}$ (but not for D_{it}).

Sequential experiments: estimation

- Yet there is a problem
- If we control $Y_{i,t-1}$, what will happen?
  Post-treatment bias for $D_{i,t-1}$ (but not for D_{it}).
- What if we don't?

Sequential experiments: estimation

- Yet there is a problem
- If we control $Y_{i,t-1}$, what will happen?
  Post-treatment bias for $D_{i,t-1}$ (but not for D_{it}).
- What if we don't?
  Omitted variable bias for $D_{i,t-1}$.

Sequential experiments: estimation

- Two approaches: structural nested mean models (SNMMs) and marginal structural models with inverse probability of treatment weighting (MSMs with IPTWs)

Sequential experiments: estimation

- Two approaches: structural nested mean models (SNMMs) and marginal structural models with inverse probability of treatment weighting (MSMs with IPTWs)
- See Blackwell (2013) and Blackwell and Glynn (2018) for more details

Sequential experiments: estimation

- Two approaches: structural nested mean models (SNMMs) and marginal structural models with inverse probability of treatment weighting (MSMs with IPTWs)
- See Blackwell (2013) and Blackwell and Glynn (2018) for more details
- The former models Y (response surface) and the latter models D (treatment assignment)

Sequential experiments: estimation

- Let's assume linearity and one lag is necessary for sequential ignorability

Sequential experiments: estimation

- Let's assume linearity and one lag is necessary for sequential ignorability
- SNMMs:
  - Regress Y_{it} on $(Y_{i,t-1}, D_{it}, D_{i,t-1})$ and get the coefficient of D_{it}, $\hat{\gamma}_0$.
  - Transform Y_{it} into $\tilde{Y}_{it} = Y_{it} - \hat{\gamma}_0 D_{it}$
  - Regress $\tilde{Y}_{it}$ on $(Y_{i,t-2}, D_{i,t-1}, D_{i,t-2})$ and get the coefficient of $D_{i,t-1}$, $\hat{\gamma}_1$.

Sequential experiments: estimation

- Let's assume linearity and one lag is necessary for sequential ignorability
- SNMMs:
  - Regress Y_{it} on $(Y_{i,t-1}, D_{it}, D_{i,t-1})$ and get the coefficient of D_{it}, $\hat{\gamma}_0$.
  - Transform Y_{it} into $\tilde{Y}_{it} = Y_{it} - \hat{\gamma}_0 D_{it}$
  - Regress $\tilde{Y}_{it}$ on $(Y_{i,t-2}, D_{i,t-1}, D_{i,t-2})$ and get the coefficient of $D_{i,t-1}$, $\hat{\gamma}_1$.
- MSMs with IPTWs: estimate the propensity score for each X_{it}, and use IPTW for ATT (Jim Robins's g-formula).

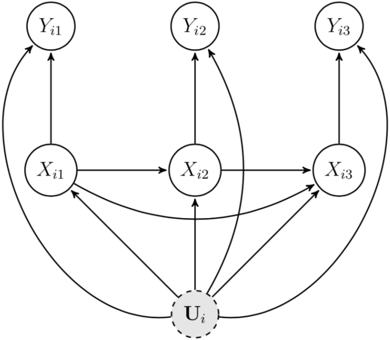
$$W_{it} = \prod_{t=1}^{t} Pr[D_{it} | D_{i,t-1}, Y_{i,t-1}]$$

The cost of allowing for unobservable confounders

- The DAG behind two-way fixed effects models

The cost of allowing for unobservable confounders

- The DAG behind two-way fixed effects models



- No feedback, no carryover, no between-unit interference

The cost of allowing for unobservable confounders

- What is the non-parametric estimator in TSCS data under the assumption of strict exogeneity?

The cost of allowing for unobservable confounders

- What is the non-parametric estimator in TSCS data under the assumption of strict exogeneity?
  Doesn't exist. Only one observation exists for unit i, period t (Imai and Kim, 2019).

The cost of allowing for unobservable confounders

- What is the non-parametric estimator in TSCS data under the assumption of strict exogeneity?
  Doesn't exist. Only one observation exists for unit i, period t (Imai and Kim, 2019).
- What if the treatment of one unit affects the outcome of others?

The cost of allowing for unobservable confounders

- What is the non-parametric estimator in TSCS data under the assumption of strict exogeneity?
  Doesn't exist. Only one observation exists for unit i, period t (Imai and Kim, 2019).
- What if the treatment of one unit affects the outcome of others?
  Wang (2020): two-way FE models lead to meaningless estimates

The cost of allowing for unobservable confounders

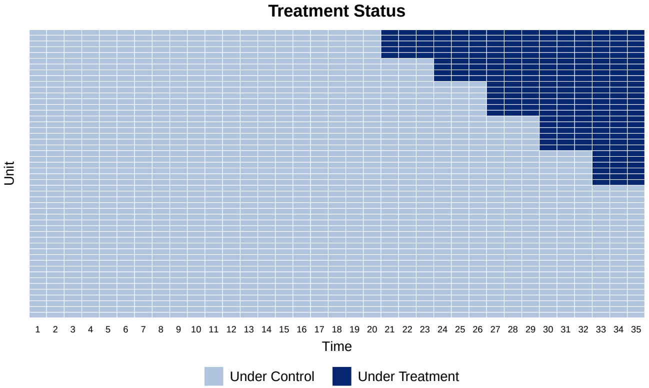
- When confounders are unobservable, we cannot estimate the propensity score any more.
- Response surface adjustment is the only possible approach.
- Either we write down a DGP for the response surface or we impose assumptions on its “trajectory”

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}

Semi-parametric models: setup

- Basic idea: predict the counterfactual of Y_{it}



Semi-parametric models: setup

- The simplest form: FEct

Semi-parametric models: setup

- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

Semi-parametric models: setup

- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

The only problem is that we cannot apply the within-estimator here (unbalanced panel).

Semi-parametric models: setup

- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

  The only problem is that we cannot apply the within-estimator here (unbalanced panel).
- We can make the error structure more complicated by replacing $\alpha_i + \zeta_t$ with $\lambda_i \mathbf{f}_t$, where $\lambda_i = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ and $\mathbf{f}_t = (f_1, f_2, \ldots, f_r)$ (interactive FE).
  How many parameters we need to estimate?

Semi-parametric models: setup

- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

  The only problem is that we cannot apply the within-estimator here (unbalanced panel).
- We can make the error structure more complicated by replacing $\alpha_i + \zeta_t$ with $\lambda_i \mathbf{f}_t$, where $\lambda_i = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ and $\mathbf{f}_t = (f_1, f_2, \ldots, f_r)$ (interactive FE).
  How many parameters we need to estimate?
- From a different perspective, we are approximating the matrix $\mathbf{Y} = \{Y_{it}\}$ with the product of two low-dimension matrices: $\mathbf{Y} = \mathbf{L} + \varepsilon$, where $\mathbf{L} = \Lambda\mathbf{F}$ (matrix completion or MC).

Semi-parametric models: setup

- The simplest form: FEct

$$Y_{it}(0) = \mu + \mathbf{X_{it}}\beta + \alpha_i + \zeta_t + \varepsilon_{it}$$

  The only problem is that we cannot apply the within-estimator here (unbalanced panel).
- We can make the error structure more complicated by replacing $\alpha_i + \zeta_t$ with $\lambda_i \mathbf{f}_t$, where $\lambda_i = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ and $\mathbf{f}_t = (f_1, f_2, \ldots, f_r)$ (interactive FE).
  How many parameters we need to estimate?
- From a different perspective, we are approximating the matrix $\mathbf{Y} = \{Y_{it}\}$ with the product of two low-dimension matrices: $\mathbf{Y} = \mathbf{L} + \varepsilon$, where $\mathbf{L} = \Lambda\mathbf{F}$ (matrix completion or MC).
  IFE and MC are equivalent when MC directly penalizes the matrix dimension r (hard-impute) rather than the magnitude of eigen values (soft-impute).

Semi-parametric models: estimation

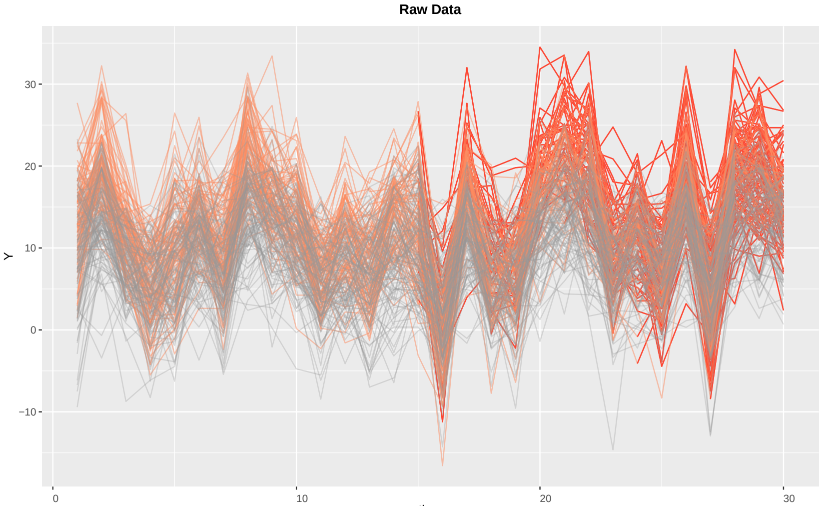
- IFE relies on the singular value decomposition (SVD).
- We start from initial parameters, estimate β via OLS, and obtain λ_i and $\mathbf{f}_t$ via SVD of the residuals.
- r is selected by cross-validation.
- MC (soft-impute) uses the direct sum decomposition: the residual at each round can be decomposed into two orthogonal parts.
- MC (soft-impute) is biased in finite samples.
- Both work well in large samples (relative performance depends on the strength of factors).
- We can add other parts to the model such as the lagged dependent variable or ensemble methods (Athey et al., 2019)

Semi-parametric models: tools

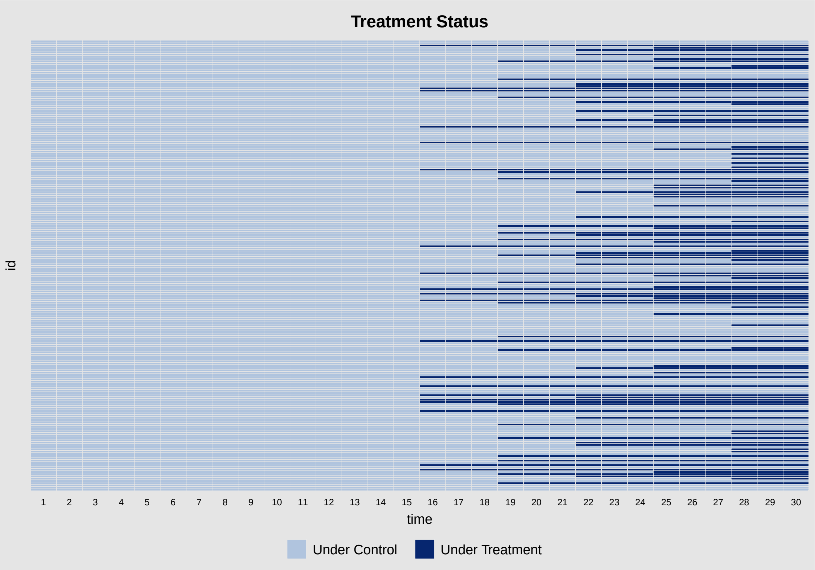
- A series of packages provided by Yiqing Xu at Stanford and his collaborators
- panelView for displaying the basic patterns
- fastplm for estimating the two-way FE models fast
- gsynth for estimating the interactive FE models
- fect to rule them all

Example

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```



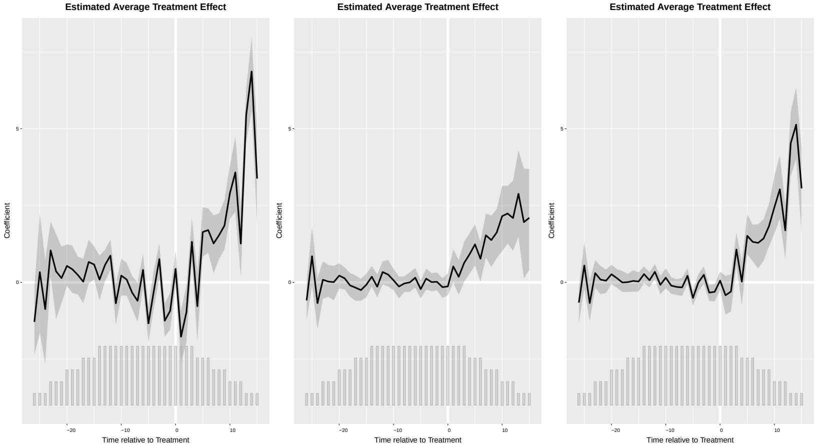
Example


Treatment Status
id
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
time
Under Control
Under Treatment

Example

```
##
##     Coef Std. Error t value Pr(>|t|) CI_lower CI_upper
## D  0.249      0.163   1.530        0   -0.070    0.569
## X1 0.999      0.041  24.432        0    0.918    1.079
## X2 2.941      0.042  69.729        0    2.858    3.024
##
## ---
##
## Residual standard error: 3.171 on 5768 degrees of freedo
## Multiple R-squared(full model): 0.800    Adjusted R-squa
## Multiple R-squared(proj model): 0.496    Adjusted R-squa
## Wald F-statistic: 1856.764 on 3 and 5768 DF, p-value: <
##
## ---
##
## Variance Type:  Robust
```

Example



Estimated Average Treatment Effect
Estimated Average Treatment Effect
Estimated Average Treatment Effect
Coefficient
5
0
−20
−10
0
10
Time relative to Treatment

Assumptions behind the semi-parametric models

- Both IFE and MC require strong exogeneity.
- It is a generalized version of "parallel trends" in DID.

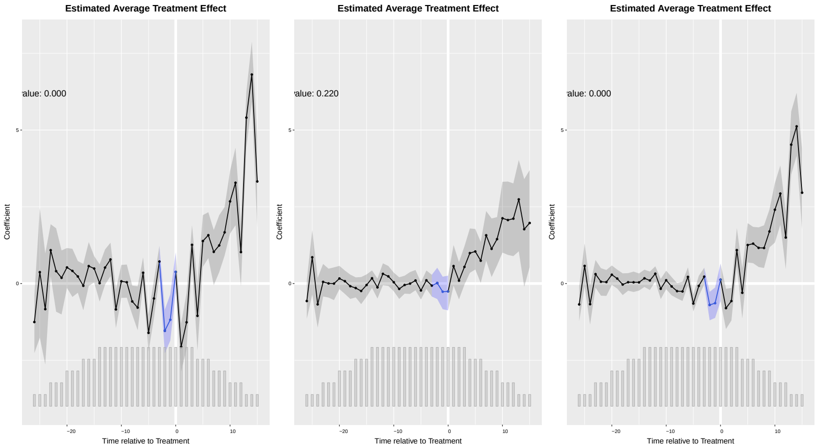
Assumptions behind the semi-parametric models

- Both IFE and MC require strong exogeneity.
- It is a generalized version of "parallel trends" in DID.
- How to test the assumption in this more general case?

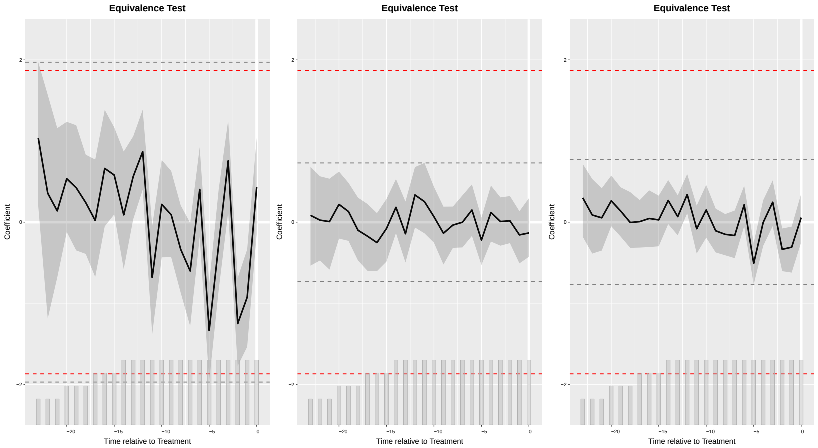
Assumptions behind the semi-parametric models

- Both IFE and MC require strong exogeneity.
- It is a generalized version of "parallel trends" in DID.
- How to test the assumption in this more general case?
- Liu, Wang, and Xu (2019): dynamic treatment effects, equivalence test and placebo test.

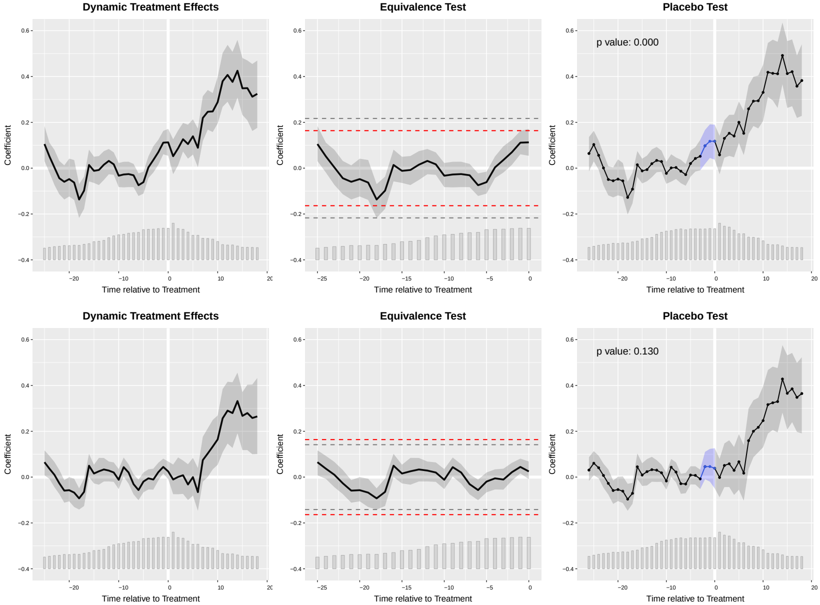
Example


Estimated Average Treatment Effect
value: 0.000
Coefficient
Time relative to Treatment
Estimated Average Treatment Effect
value: 0.220
Coefficient
Time relative to Treatment
Estimated Average Treatment Effect
value: 0.000
Coefficient
Time relative to Treatment

Example


Equivalence Test
Equivalence Test
Equivalence Test
Coefficient
Coefficient
Coefficient
2
0
−2
−20
−15
−10
−5
0
Time relative to Treatment
Time relative to Treatment
Time relative to Treatment

Example: Tomz et al. (2007)



Trajectory balancing

- One common assumption: conditional parallel trends

$$E[Y_{it}(0) - Y_{i,t-1}(0) | D_{it} = 1, X_{it}] = E[Y_{it}(0) - Y_{i,t-1}(0) | D_{it} = 0, X_{it}]$$

- Then $E[Y_{it}(1) - Y_{i,t-1}(0) | D_{it} = 1, X_{it}] = E[Y_{it}(1) | D_{it} = 1, X_{it}] - E[Y_{i,t-1}(0) | D_{it} = 1, X_{it}] - E[Y_{it}(0) | D_{it} = 0, X_{it}] + E[Y_{i,t-1}(0) | D_{it} = 0, X_{it}]$
- Synthetic control, Semi-parametric DID (Abadie, 2005), kernel balancing (Hazlett and Xu; 2017), panel matching (Imai, Kim, and Wang; 2018)
- One can use weighting to adjust the difference in trends among groups.
- Remember that this is not IPW and we can only get ATT not ATE.

A new hope

- Arkhangelsky and Imbens (2019)
- Even if we allow for the existence of unobservable confounders, it may still be possible to estimate the propensity scores.
- We need stronger assumptions on the assignment mechanism.
- For example, when the true propensity score has the form:

$$E\left[D_{it}|\alpha_i\right] = \frac{exp(\alpha_i + \xi_t)}{1 + exp(\alpha_i + \xi_t)}$$

  We have $\mathbf{D}_i^T \perp \mathbf{Y}_{jt}(\mathbf{D}^T)|\frac{1}{T}\sum_{t=1}^{T} D_{it}$.
- They show that there exists a doubly-robust estimator under these assumptions.